Matej Sembera

Sonnenallee 24, Vienna, 1220 | semb.mato@gmail.com | +43 664 217 77 41 | github.com/matej897

SUMMARY

I am a Data Science student at the University of Vienna with one year of experience as a Data Science intern. I have solid knowledge in data analysis, business intelligence, AI, and data engineering.

I am eager to apply my skills in real-world projects while continuing my studies. As an EU citizen, I have unrestricted work authorization. I also hold a Goethe Certificate at the C1 level in German language proficiency.

EDUCATION

University of Vienna, Master's in Data Science   Sept 2024 – Present

- Coursework: Data Ethics, Introduction to Machine Learning, Visual and Exploratory Data Analysis.

Comenius University Bratislava, Bachelor's in Financial and Economic   Sept 2021 – June 2024
Mathematics

- Coursework: Statistics, Theoretical Mathematics, Economics, Accounting, and Actuarial Mathematics.

WORK EXPERIENCE

Data Science Intern, Allekommen.at - Vienna, Austria   June 2023 – June 2024

- **Created dashboards and data visualizations (Power BI, Tableau)**

- **Data analysis & reporting (Excel, Python, R, SQL):** Structured, cleaned, and analyzed large datasets. Created and optimized reports and dashboards for management. Conducted ad-hoc analyses to quickly address urgent business questions.

- **AI implementation (Python, NLP, LLM):** Contributed to a team that integrated a Natural Language Processing model (OpenAI), enabling the system to return relevant results even in cases of typos (e.g., "Hiarcut" instead of "Haircut"). This **significantly reduced user drop-off rates**.

- **Development of a SQL-based search algorithm (Python, SQL, geodata):** Built a new GeoSQL search algorithm that is **10 times** faster than the previous version and supports high user traffic.

- **Built ETL pipelines (Python, Linux):** Automated data collection via web scraping, processed data using Python and Linux shell commands, and stored it in a PostgreSQL database **(approximately 2,000 service providers uploaded)**.

- **Optimized SQL queries and development in PostgreSQL (SQL):** Analyzed and optimized SQL queries, developed functions for the search algorithm, and removed duplicate data.

- **Reference:** Michael Pisnyachevskiy, michael@allekommen.at

PROJECTS

Bachelor Thesis: Machine Learning for Stock Index Prediction (Grade: A)   Link to GitHub

- In my bachelor thesis, I used machine learning to predict next-day stock market movements and traded based on these predictions. On the test dataset (Nov 23, 2021 – Jan 24, 2024), the model outperformed the S&P 500 by **12.8%**.

Course Project: Information Management & System Engineering   Link to GitHub

- In this course project, we built a bicycle shop website using JavaScript and TypeScript, containerized it with Docker, and migrated the database from PostgreSQL to MongoDB using .NET.

CERTIFICATES

Associate Data Scientist in Python   Link to GitHub

- Proficient in Python and libraries such as Pandas, Matplotlib, Seaborn, NumPy, TensorFlow, Scikit-learn, and PyTorch.

SQL Fundamentals   Link to GitHub

IBM Data Engineering Professional Certificate   Ongoing
- Built multiple ETL pipelines using Python and Linux shell commands to extract data from XML, JSON files, and websites. Gained foundational knowledge of tools like Airflow, Kafka, Hadoop, Apache Spark, and AWS.

SKILLS

Languages: German (C1, Link to Goethe Certificate C1), English (C1, Link to IELTS), Slovak (C2)
Soft Skills: Analytical thinking, high sense of responsibility, structured approach to complex problems.
Programming Languages: Python, R, SQL.
Tools: Excel, Power BI, Tableau, GitHub, Docker, Power Query, Airflow, Kafka, Hadoop, Apache Spark.
Databases: PostgreSQL, MongoDB, MySQL.
Data Formats: XML, JSON.
Cloud Platforms: AWS.
MS Office: Good proficiency.